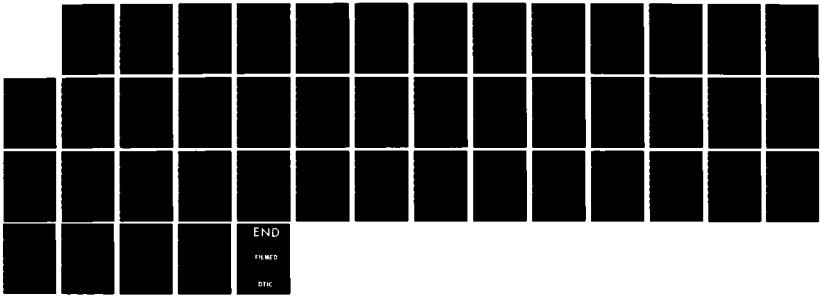
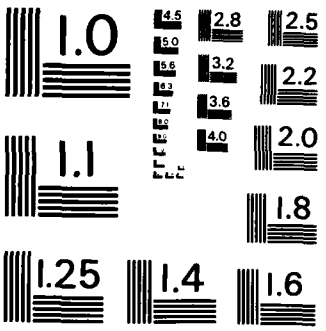
AD-A159 570 WORK STATUS CHOICE AND THE DISTRIBUTION OF FAMILY EARNINGS(U) RAND CORP SANTA MONICA CA N U HAQUE NOV 84 RAND/P-7037 1/1

UNCLASSIFIED F/G 5/11 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(2)

WORK STATUS CHOICE AND THE DISTRIBUTION OF FAMILY EARNINGS

Nadeem Ul Haque

November 1984

P-7037

ACKNOWLEDGMENTS

For helpful comments, the author is grat
Julie DaVanzo, Dennis De Tray, Peter Kostiuk,
Schultz, Robert Topel and the members of the
Applications of Economics at the University
omissions, of course, remain the responsibi

INTRODUCTION

Analyses of the distribution of earnings are concerned normally with the measurement and explanation of observed inequality in the personal earnings of individuals (see Mincer [1974], Psacharopoulos [1977] and Psacharopoulos and Layard [1979]). Such analyses often seek to analyze the returns to human capital as measured by age and experience while using other individual characteristics as control variables. As a consequence, issues of individual self-selection into occupations which in themselves affect the eventual distribution of earnings are often not explicitly analyzed. Furthermore, observations of the self-employed segment of the labor force are often excluded from the analysis because of the difficulty of separating the wage component from that of the return to capital. Such a practice in the context of industrial economies where the self-employed are a very small proportion of the labor force is somewhat defensible. However, in developing economies where, as has been noted in earlier studies such as Chiswick [1977], Haque [1977] and Wong [1981], the self-employed can constitute as much as up to a third of the labor force, basing the analysis of the distribution of earnings solely on the employed segment of the labor force can lead to serious biases.

An explicit analysis of the endogeneity of the work-status decision, i.e., that between employment and self-employment, could contribute to a better understanding of labor markets in less developed countries and hence help in the making of policy. In this connection, we are able to study empirically whether the self-employed workers are those who would not do well in the formal labor markets, as the labor market segmentation hypothesis implies. Moreover, insofar as the self-employed workers are representative of the "informal" sector as defined in the segmentation literature (see Mazumdar [1981]), they should be characterized by relatively lower earnings, greater inequality, and both lower education in terms of years of formal schooling and lower returns to such schooling. However, a competing hypothesis would suggest that individuals, based on their particular characteristics and conditions

prevailing in the market, self-select themselves into a particular work status. Consequently, earnings of an individual in the employment of his choice should be higher than what he could obtain in alternative employment. A better understanding of the empirical validity of th two hypotheses would be helpful in framing policy, since the two have obviously different policy implications, with the former suggesting a somewhat more interventionist approach.

Since the individual has most often been taken as the unit of analysis in research in this area, the effects of interdependent family decision-making have not received much attention. However, from a welfare point of view, a case can be made for studying the distribution of earnings at the level of the family, given that individual consumption and investment decisions are taken jointly in a family setting. The predication of individual decisions on family circumstance is illustrated, quite well, in recent research on female labor supply behavior (see Heckman [1979] and Smith [1979]). These studies have observed that husband's income and the number of young children are germane to the female decision to enter the labor force. Furthermore, as noted by Smith [1979], Layard and Zabalza [1979] and Gronau [1981], the employment of the wife tended to produce an equalizing effect on the observed distribution of family earnings. These studies, however, dealt mainly with developed economies where the family structure was nuclear. Our sample, on the other hand, while being drawn from a developing economy, reports a large proportion of nonnuclear families and many multiple-earner households. Moreover, since female labor force participation is very low, the secondary earners in the household are mainly male nonnuclear members of the household. Thus, to study the distribution of family earnings, both the choice of family structure and the earnings behavior of the nonnuclear family members have to be analyzed.

The paper is divided into six sections. The first section describes features of the data upon which the analysis is based. Section II estimates the rates of return of education in the Becker-Mincer human capital framework and notes the unusually low rates for self-employment. As a consequence of this differential effect of education in the two employment status categories, work status choice is

treated as an endogenous variable in the study of earnings behavior in Section III. Section IV considers family earnings variations across work status choice and shows that employed families reveal more widely dispersed earnings than families where the head of household is self-employed. At lower levels of income however, the opposite holds; self-employed family earnings are more dispersed. Section V treats the question of multiple-earner households. Correlations in potential earnings within a household are shown to be larger than correlations in observed earning; the corrected dispersion in the distribution of family earnings would therefore be larger than observed. Section VI examines the relationships among family wealth, family structure, and the probability of finding a secondary earner in a household. Choice of family structure is based on economic consideration and poorer households are more likely to release secondary earners into the labor market.

I. THE DATA

The data upon which the analysis is based derives from a survey of Rawalpindi, the fifth largest city in Pakistan, with a population of 716,761 in 1977 at the time of the survey[1] and historically an important regional metropolis and administrative center. It was the nation's capital during the 1960s and today is only ten miles from the current capital, Islamabad. Though little manufacturing activity is located in the city, the development of Islamabad on the outskirts of Rawalpindi has allowed an expansion of wholesale trade and construction.

In August and September of 1977, the Pakistan Institute of Development Economics undertook a survey to collect socioeconomic information on 2000 Rawalpindi households. The small sample size and the restriction to one city are the result of a limited budget. In the two-stage sampling procedure, the household clusters first were randomly selected from a 400 cluster sampling frame (a cluster being a group of approximately 263 households. At the second stage a fixed proportion of households per cluster (2000/16 = x) were selected from the clusterwise address list of all structured and semi-structured dwelling. A fairly detailed questionnaire was developed, pretested, and revised.[2] The revised questionnaire was completed for each household through an interview with the head of household or the oldest household member. Numerous checks were devised to ensure a proper implementation of the survey and accuracy of the information.

A description of the labor force and family characteristics of the data set are set out in Table 1. Among the more striking features of the data are the extremely low level of female labor force participation, (6.81 percent) the large proportion of self-employed in

---

[1] The population figure is a projection of the 1972 census estimate at a 3.2 percent annual growth rate.

[2] A major pretest for the 1977 Rawalpindi survey was the smaller 1975 survey of 1000 households (see Hamdani [1977] and Haque [1977]). Attempts were made in the 1977 survey to go back to the same 1000 households of the earlier survey and at the same time collect information on 1000 more.

the labor force, (40 percent), and a significant proportion of nonnuclear families in the sample (36.7 percent). The first finding probably is explained by the population's cultural and religious norms. The proportion of self-employed seems high even when one takes into consideration that the sample is in an urban area of a developing economy. Compared with other samples from similar economies, the reported proportion of self-employed in Rawalpindi may be high. Wong [1981], for example, reports 16 percent self-employed in the total labor force in Hong Kong.[3]

The sample reports a feature which has been observed quite commonly among developing economies--a large proportion of observed families reporting a nonnuclear or an extended structure. This information can be helpful to the analysis of the choice of family structure and the interaction of this choice and other forms of household decisionmaking, such as individual labor supply. In this view, valuable insights into family decisionmaking, an area of continuing interest (see Becker [1981]) in economics, can be gained. Given the low level of female labor force participation, about 38.5 percent of all earners are either children of the household, who have chosen gainful employment over the alternative of furthering their schooling, if any, or extended family members.

[3] A probable explanation is the large rural-urban migration into Rawalpindi from a population hinterland. Hong Kong does not have such an immigration.

Table 1

THE DATA

I.	Description of the labor force:	
	Total number of individuals in labor force	3,287
	% female in labor force	6.81
	% of labor force unemployed	15.97
II.	Composition of the gainfully employed labor force[a]:	
	Occupational choice	
	% employed	53
	% self-employed without employees	37
	% self-employed with employees	3
	% apprentices and casual employees	7
	Relationship	
	% head of household (HH)	57.3
	% wife of HH	4.2
	% parent of HH or wife of HH	4.6
	% children of HH	19.1
	% sibling of HH or wife	8.3
	% others	6.5
III.	Description of family structure	
	Average family size (number)	6.16
	% nuclear families[b]	63.9
	% extended families[c]	36.1

[a]In this category the estimates are percentage of those who are currently reporting positive earnings.

[b]Nuclear families are defined as parents and children under 20.

[c]Extended families are defined as nonnuclear.

II. SCHOOLING AND EARNINGS

In the Becker-Mincer model of human capital, a man with no schooling or training is assumed to earn E_0 in each year of his life. Each year's education raises earnings by a constant percentage amount γ, which is the rate of return to schooling. Denoting by E_s the level of earnings of a man with s years of schooling but no training we have

$$E_s = E_0 e^{\gamma s}. \qquad (1)$$

Table 2

DEFINITION OF VARIABLES"

Variables	Definition
CURHSOWN	1 if HH owns house the family lives in, 0 otherwise.
CURPACCA	1 if house HH lives in is made of concrete, 0 otherwise.
E	1 if individual self-employed.
EXP	Years of individual work experience.
FEMALE	1 if female, 0 otherwise.
FCLAS	No. of years of completed education by HH's father.
GRAD	Years of completed education.
HE	HH employment status, 1 if HH self-employed.
HEDAGE	HH age in years.
HEDGRAD	HH years of completed education.
HH	Head of household.
ln Y	Natural log of earnings.
MCLAS	No. of years of completed education by HH's mother.
MIG	1 if HH migrated to city less than 10 years ago, 0 otherwise.
S	Years of schooling.
URBAN	1 if HH spent childhood in urban environment, 0 otherwise.
XTEN	1 if extended family (nonnuclear), 0 otherwise.
YRSWKED	Total years of work experience.
YRSATJOB	Number of years at current job.
PUCCA	1 if house HH lives in is made of concrete, 0 otherwise
OWNHS	1 if HH father owned house, 0 otherwise.
PUCHS	1 if HH parents owned a house made of concrete, 0 otherwise.
RUNWAT	1 if there was running water in HH's father's house, 0 otherwise.

Such E_s-constant earnings profiles would be observed if jobs did not add to individual's market skills. Training on the job and learning by doing do contribute to such development. Individuals are therefore willing to pay for the acquisition of such skills by forgoing part of their potential earnings. If E_t is the potential earnings of the individual in each period and $k_t E_t$ his investment in post-school training (k_t being the net investment ratio), we have[4]

$$\ell n\ E_t = \ell n\ E_0 + \gamma S + \int_0^t \gamma k_z dz. \qquad (2)$$

Observed earnings Y_t are therefore the difference between potential earnings and the value of the post-school investment:

$$\ell n\ Y_t = \ell n\ E_t + \ell n\ (1 - k_t). \qquad (3)$$

Given a functional form for the time path of k_t, we can obtain an estimable form for our earnings function. The most commonly used form, following Mincer [1974], is linear

$$k_t = k_0 - bt. \qquad (4)$$

The net investment ratio and therefore the value of the post-school investment declines over time. Substituting (2) and (4) into (3) we are able to obtain the Becker-Mincer human capital earnings function.[5]

$$\ell n\ Y_t = \ell n E_0 + k_0\ (1 + \frac{k_0}{2}) + \gamma S + [\gamma k_0 + b(1 + k_0)]X$$

$$- (\frac{\gamma b}{2} + \frac{b^2}{2})X^2 + u_t. \qquad (5)$$

---

[4] It has been assumed here that both schooling and post-schooling investment bear the same rate of return, γ. This assumption is necessary for the identification of k_0 and b.

[5] After integration and adding on an error term, we get to equation 5.

The returns to schooling estimated in the earnings function over the whole sample are averages of the effects of schooling per se and the effects of training during working life. Investment in post-school training typically takes place early in the life cycle in the form of forgone earnings from potential earnings, while benefits accrue later in life. To obtain the true rate of return to schooling, the method used by Mincer isolates the effect of experience, relying on the point of "overtaking," defined to be the point where potential earnings, free of training costs and returns, are equal to observed earnings. If the overtaking year is denoted by ℓ, we have

$$\ell n\ Y_{\ell} = \ell nE_0 + \gamma S + U_{\ell} \qquad (6)$$

By assuming a common overtaking year for all individuals, the overtaking point can be identified as one where the explanatory power of the regression of log of earnings on schooling is at its maximum. We would now be able to identify γ by estimating (6) for the overtaking set alone.

By this method the overtaking set was determined to be individuals with 3-5 years of work experience. The regression results for this set are presented in Table 3. Because a large proportion of the sample reported a self-employed work status (37 percent), separate estimates were made for that group as well. For the full sample, returns to schooling are estimated at 8.9 percent. For employees the estimate is higher still at 11.2 percent, but much lower for the self-employed, standing at 2.5 percent only. The extremely small estimate for the return on schooling for the self-employed is striking. For Hong Kong (see Wong [1981]) γ was estimated to be similar for both the self-employed and the employed.

The functional form that we derived earlier and the estimates which we have just discussed, assume a constant return to each additional year of schooling. In order to go beyond constraining returns-to-schooling to such a linear pattern, equations with a quadratic functional form for returns-to-schooling were also estimated and are also presented in Table 3. The results of these estimations indicate increasing returns to

Table 3

OVERTAKING YEAR REGRESSIONS FOR MALES
(t-RATIOS IN PARENTHESES)

Variables	Employees		Self-Employed		Full Sample	
Constant	4.84	4.93	5.83	5.85	5.12	5.20
	(63.23)	(57.10)	(38.47)	(34.38)	(70.15)	(62.94)
S	.112	.058	.025	.012	.089	.042
	(13.67)	(2.22)	(1.21)	(.20)	(10.99)	(1.64)
S^2		.004		.001		.003
		(2.14)		(.21)		(1.95)
R^2	.461	.47	.260	.022	.294	.303
σ^2	.362	.356	.409	.523	.454	.450
mean(s)	6.88		4.02		5.86	
σ^2(s)	23.10		17.63		23.03	
mean(ln y)	5.92		6.20		6.02	
σ^2(ln y)	.569		.416		.532	
mean (ln y*)[a]	5.73		5.98		5.79	
σ^2(ln y*)	.667		.420		.641	
R^2	.364		.017		.149	
mean (exp)	15.09		23.17		17.97	
σ^2(exp)	149.28		203.802		183.61	
Number of Observations	221		71		292	

[a] ln y* denotes log earnings of men in the overtaking experience years.

additional years of schooling. Estimates for completed levels of education are 7.2 percent for completed primary schooling, i.e. 5 years of schooling; 10.2 percent for a high school diploma, i.e. 10 years of schooling; and 13.8 percent for a university graduate, i.e. 16 years of schooling. For the employees at the same levels of schooling the estimates are 9.8 percent, 13.8 percent and 18.6 percent respectively.

earner, 10.7 percent two secondary earners, and about 2 percent three or more secondary earners. Thus, for our analysis the concentration is on the influence of secondary-earner incomes on family earnings.[10]

Table 8 presents for both categories of work status the variances of the logarithm of earnings. Families are further stratified by age of HH and the presence or absence of a secondary earner. Measured by the variance of the logarithm of earnings, HH earnings are more unequally distributed for the self-employed while family earnings are more dispersed for the employed. Within employment categories, family earnings are more unequally distributed than HH earnings for all employed age groups, whereas only for the younger self-employed age groups (HH age less than 40) is this true. For the older self-employed groups, family earnings are less dispersed than HH earnings. For families with no secondary earners, total family earnings are everywhere (except for the oldest self-employed group) more spread out than for families with secondary earners. The finding of greater dispersion in

Table 8

INEQUALITY IN EARNINGS AS MEASURED BY VARIANCES IN LOG EARNINGS

	Family Earnings			HH Earnings	
Age Employed	All	With Secondary Earner	No Secondary Earner	All	With Secondary Earner
Employed					
21-30	.3393	.1978	.2133	.1965	.1715
31-40	.2562	.2059	.1938	.2021	.2327
41-50	.3370	.2956	.3054	.2976	.2489
51-60	.3669	.2224	.2317	.2712	.2978
Self-Employed					
21-30	.3245	.1588	.3194	.2807	.1277
31-40	.2245	.1001	.2265	.2121	.094
41-50	.2701	.2330	.2405	.3094	.4157
51-60	.3505	.4277	.2630	.4498	.5585

[10] For this analysis the sample is restricted to a married spouse, present families only, with an earning head of household.

$$C^2 = \alpha_1^2 C^2 + (1 - \alpha)^2 C_2^2 + 2\alpha(1 - \alpha)\rho C_1 C_2. \quad (11)$$

Here C is the coefficient of variation in the distribution of family income, and C_1 and C_2 are the coefficients of variation in husbands' and wives' incomes respectively, ρ is the correlation coefficient between spouse earnings and α is the share of husband's income in total earnings--$\alpha = \frac{\mu_1}{\mu_1+\mu_2}$, where μ_1 and μ_2 are the mean earnings of husbands and wives respectively. Define $E_1 = C/C_1$ and $\theta = C_2/C_1$. Equation (11) can now be rewritten as

$$E^2 = \alpha^2 + (1 - \alpha)^2\theta^2 + 2\alpha(1 - \alpha)\rho\theta.$$

Family income inequality is therefore reduced by the accounting of wives' earnings only if $E < 1$, or

$$C_1 > \frac{C_2}{(1 - \alpha)} \left[-\rho\alpha + \left\{1 - (1 - \rho^2)\alpha^2\right\}^{\frac{1}{2}}\right].$$

This inequality is more likely to hold as α increases and it is less likely to hold as ρ or θ increases. Further complications can be introduced by the explicit dependence of C_2, α, and ρ on female labor force participation decisions and female characteristics. The marginal effect of an additional worker's earnings on combined family earnings is, therefore, indeterminable.

The study of the distribution of earnings in our data is of interest because of two characteristics revealed by the sample. First, the extremely low labor force participation rates of women mean that families depend for their market purchases on the incomes of male members of the household alone.[9] Second, many families report multiple earners per household--the earners other than the head of household being sons, siblings, or fathers of either the head of household (HH) or his spouse. Some 27.7 percent of the households report one secondary

[9] This statement and much of the analysis that follows ignores the important contribution of women in the household production, mainly because of data considerations.

IV. THE DISTRIBUTION OF FAMILY EARNINGS

Studies of the distribution of earnings frequently use the individual for their unit of analysis, even though from a welfare point of view attention should be focused on the family. Much of an individual's life is spent in the family grouping where decisions are either taken jointly or predicated on the family structure and wealth. Thus individual consumption, investment, labor supply, and other decisions of interest to welfare analysts are at the very least dependent on family circumstances. From a welfare point of view, the concern with the distribution of resources among the members of society is in the main answered by how resources are shared among the families of that society. Individual work decisions, in particular the observation of multiple earners in poorer households, could result in a less skewed distribution of income. Intertemporal individual labor market participation decisions could protect the family from the vicissitudes of fortune.

The study of the distribution of family earnings is fairly new and has been conducted for the United States by Smith [1979], for the United Kingdom by Layard and Zabalza [1979], and for Israel by Gronau [1981]. The concern of each of these studies has been with measuring the influence of the wife's earnings on the distribution of family incomes. The conclusion in all cases was that the inclusion of female earnings tended to equalize observed family income distribution. There is no a priori reason to expect such a result, however, because family income distribution in a nuclear context depends on the interaction of female labor force participation and labor supply decisions, assortative mating and individual market characteristics. For example, positive correlation between wages of husbands and wives resulting say from positive assortative mating might be balanced off by negatively correlated labor supply decisions to leave the family income distribution unaffected. The interaction of these various determinants of family income distribution is illustrated by examining their effect on one commonly used measure of dispersion, say the coefficient of variation. The coefficient of variation as a weighted average of individual coefficients of variations with the weights being individual contributions to family income can be written as

As before, the educational and experience variables are significant in explaining earnings variation. Returns to education and experience continue to be smaller among the self-employed. The current job experience variable measuring the job-specific human capital acquisition is, as expected, positive and also fairly constant across employment categories.

Family educational background as measured by mother's and father's education, is negatively related to individual earnings. Of the wealth variables, once again only the availability of running water in parental house is significant and positive probably because it is a sharp wealth proxy. Because ownership of house and land by parents gives little information on the quality of the asset, it is not effective in our statistical tests. The current wealth variables have the expected positive signs but the ownership of the current dwelling is insignificant. CURPUCCA, which selects individuals who own a current concrete house, now controls for the quality of the owned asset. Consequently, it has the right sign and is marginally significant. A motive for including the current wealth variables was to attempt to isolate the effect of capital on the earnings of the self-employed. The CURPUCCA coefficient is larger for the self-employed category.

An individual who has been in the city more than seven years has for the purposes of this analysis been regarded as a non-migrant. Table 7 thus indicates if migrants are able to find employment, they earn as much as the average employee. This could be a product of the possible large weighting of the government employees among the employee subsample and the rigidity of the government employment structure. Self-employed migrants earn significantly less than their employed counterparts, however. The segmentation conjecture suggests such a result, arguing that the recent migrants would find it harder to find employment owing to their lack of location-specific capital. To acquire knowledge of the local terrain they spend time in the lower end of the self-employed distribution. Furthermore, an urban childhood environment does lead to a higher wage in both sectors, indicating returns to location-specific knowledge.

Table 7

EARNINGS FUNCTIONS CORRECTED FOR SELECTIVITY BIAS:
MALE HEADS OF HOUSEHOLD*

	Employed		Self-Employed	
Variable	Selectivity Corrected	Uncorrected	Selectivity Corrected	Uncorrected
Education:				
S	-.0216	-.0217	-.0141	-.0161
	(2.46701)	(2.4732)	(.7761)	(.8633)
S^2	.0063	.0062	.0034	.0031
	(9.7637)	(9.5811)	(2.0421)	(2.0838)
Experience:				
EXP	.0300	.0330	.0146	.0206
	(6.3335)	(7.4221)	(1.6956)	(2.4540)
EXP^2	-.0004	-.0005	-.0003	-.0005
	(3.4129)	(5.7611)	(1.3909)	(3.3557)
YRSAT JOB	.0050	.0048	.0045	.0044
	(2.8160)	(2.7393)	(1.9570)	(1.9200)
Educational Background:				
FCLAS	-.0172	-.0015	-.0438	-.0045
	(1.8060)	(.4152)	(2.4912)	(.4925)
MCLAS	-.0056	-.0018	-.0327	-.0211
	(.7814)	(.2628)	(.5933)	(1.0496)
Wealth Background:				
PUCCA	.0190	.0079	.0676	.0552
	(.6020)	(.2554)	(1.2568)	(1.0251)
OWNHS	-.0121	-.0277	.0691	.0453
	(.3130)	(.7307)	(1.0706)	(.7055)
RUNWAT	.6147	.1558	.1949	.1566
	(3.5756)	(3.3989)	(2.1312)	(1.7250)
OWNL	-.1109	-.0328	-.1145	.0504
	(2.1124)	(1.1282)	(1.3843)	(.9412)
Current Wealth				
CURHSOWN	.0196	.0167	.0013	-.0198
	(.6709)	(.5708)	(.0253)	(.3825)
CURPUCCA	.0793	.0809	.1273	.1204
	(1.6829)	(1.7179)	(1.6407)	(1.5441)
Other				
URBAN	.0826	.0301	.1061	-.0149
	(1.7988)	(.8524)	(1.3862)	(.2430)
MIG	.0327	.0349	-.2538	-.2563
	(.8774)	(.9343)	(2.7808)	(2.7933)
λ_E	.4020			
	(1.7855)			
λ_S			.7868	
			(2.6050)	
Constant	5.6494	5.3412	5.2132	5.8725
	(30.3375)	(76.3712)	(18.0992)	(42.4664)
R^2	.4229	.4206	.1013	.0898
σ^2	.1431	.1435	.3250	.3285
Obs.	826		543	

*t-ratios in brackets.

young resulting from economic growth. Father's schooling and mother's schooling both affect the probability of employment positively, although the latter is both low in value and statistically insignificant, which is not too surprising, given low levels of female schooling and labor force participation. Education levels are positively correlated across generations, and the effect of education is strong among the employed. The positive effect of parental education indicates the increased return to education in employment. Contrary to expectation, the sign for the environment variable is negative. An urban childhood reflects the advantages of location-specific knowledge and should enhance the probability of the individual's entry into the coveted employment sector. Of the wealth variable only the ownership of land is significant and positive, lending some support to the segmentation conjecture. The weak effect of other wealth variables is understandable, given the rather weak proxy variables that we have.

Using the results of the probit, the λ_is were constructed and (10) estimated. Table 7 presents the results for regressions corrected and uncorrected for selectivity bias. The significance of the coefficients of both λ_E and λ_S confirms the presence of selectivity in our sample.[8] Positive values of both coefficients indicate positive selectivity for the self-employed and negative selectivity for the employed, which contradicts the segmentation hypothesis. In Rawalpindi, therefore, the observed self-employed earn more than a hypothetical population mean under random selection, whereas the observed employed earn less than such a mean. In view of the proximity of Rawalpindi to the nation's capital, probably most of the employed in the sample are government employees, even though our sample does not distinguish them from others. The statistical failure of the comparative advantage hypothesis is probably due to the government policy of expanding employment at the cost of real wage growth. It will be interesting to compare evidence on the occupational choice question with that in a national sample now becoming available.

---

[8] The human capital variables, schooling and experience, show little movement across the regressions corrected and uncorrected for selectivity, while the background, current wealth, and childhood environment variables are affected by the estimation procedure in terms of both the size of the coefficients and their levels of significance.

Table 6

MAXIMUM LIKELIHOOD ESTIMATES OF PROBABILITY OF EMPLOYMENT FOR MALE HEADS OF HOUSEHOLD

Variable	Coefficient	Asymptotic t-Ratio
Personal Characteristics		
HEDAGE	.0495	2.2798
$HEDAGE^2$	-.0008	3.0494
Childhood Environment		
URBAN	-.2492	2.7656
Educational Background		
FCLAS	.0724	7.1290
MCLAS	.0221	1.0279
Wealth Background		
OWNHS	-.0095	.0788
PUCCA	.0785	.4771
PUCHS	-.1334	.7345
OWNL	.3330	4.3343
Constant	.6889	1.5804
- 2 Log	843.16	

an identifying variable, excluded from the probit and included in the wage equation.[7]

The results indicate that after age 31, the probability of choosing self-employment increases with each succeeding year. Because the self-employed are an older group of people, this result may be only a cohort effect--a consequence of increased employment opportunities for the

[7] It might be noted here that "background" variables too were used as identifying variables and education included in both the probit and the earnings function. In this estimation the coefficient of education in the probit was insignificant.

$$\ell ny_i = X_i\beta_i + \frac{\delta_{ie}}{\delta_{ee}}\lambda_i + v_i \quad i = E,S. \qquad (10)$$

The V_is are now random disturbances with zero means, and the δ_{ie} is the covariance of U_i (i = E,S) with e the disturbance term in (9).

Coefficients of the λ_i are in themselves of interest. Consistent with the theory of occupational choice would be positive self-selection, or individual self-selection on the basis of comparative advantage. Observed earnings in each category should be higher than the average in a hypothetical random assignment. Since λ_E is negative and λ_S positive, this hypothesis would imply that their coefficients should have the same signs.

For less developed countries, the hypothesis of labor market segmentation deals with the simultaneous existence of low and high wage sectors with barriers to entry preventing the market from working in the direction of wage equalization (see Mazumdar [1981]). The protective barriers could be institutional, such as unionization and government legislation, and the higher wages could be in the form of fringe benefits, job security, and paid vacations, etc. The theory is weak and ill-defined in both content and statistical implications. Empirical definitions of the two sectors are in themselves a primary problem. The self-employed and the employed are often regarded as natural definitions of low and high wage sectors. In Rawalpindi, however, the self-employed earn on an average more per month than the employed, with incomes distributed slightly more equally, as measured by the log variance (see Table 4). The employed take paid vacations and work shorter hours, however, thus earning a higher hourly wage rate. The dual labor markets conjecture would imply a negative coefficient for both λ_E and λ_S, indicating that although employed positively self-select, the self-employed are unable to do so owing to some unobserved market imperfection.

The probit estimates (9) are presented in Table 6. The estimated equation includes only background variables. Because education is assumed exogenous to the decision of work status choice, it is used as an identifying variable, excluded from the probit and included in the

Estimating the earnings function while ignoring the endogeneity of work-status choice will lead to biased and inefficient results. The correct estimation procedure is to take into account the qualitative dependent variable in a simultaneous equation framework. Such simultaneous equation models with limited dependent variables have been used for the estimation of earnings functions for samples containing both unionized and nonunionized workers (Lee [1978]). Rosen and Willis [1979] used a similar model to examine the simultaneity of the decision to go to college and the returns to college education.

Estimation procedures have been dealt with extensively in Lee [1978] and Heckman [1979]. The bias in the estimation of Eq. (7) by OLS is a consequence of the truncated distribution of the error term resulting from a self-selected sample. A convenient two-step procedure due to Heckman permits correcting for this bias under the assumption of joint normality in the distribution of error terms. Rewrite Eq. (8) as

$$I = Z\gamma + e. \tag{9}$$

The first step involves the consistent estimation of (9) by maximum likelihood or weighted nonlinear least squares. Given this consistent estimate of γ, the conditional means of the error terms in the employment equation, given that we observe the individual in that state can now be constructed,

$$\lambda_E = \frac{-f(Z\hat{\gamma})}{F(Z\hat{\gamma})}.$$

Here the circumflex (^) indicates an estimate, and $f(\cdot)$ and $F(\cdot)$ are the density and the distribution function of the standard normal variable. Similarly, the conditional mean of the error term in the self-employment equation, given the observation of that individual in self-employment, is

$$\lambda_s = \frac{f(Z\hat{\gamma})}{1 - F(Z\hat{\gamma})}$$

Now Eq. (7) can be written as:

The employee subsample estimates compare favorably with those of other countries, and the full sample estimates seem to be reflecting the poor fit of the human capital model to the self-employed subsample. In later sections we will pursue the question of occupational choice and its effect on earnings further. On the issue of generalizability of results to the whole country, two peculiarities of the data set are that (1) Rawalpindi is a close neighbor of Islamabad, so the largest employer in the city is the government, and (2), the urban nature of our sample may bias the rates of return upward, given infrastructural considerations in the rural areas.

III. WORK-STATUS CHOICE AND EARNINGS

In this section, earnings functions will be estimated treating the choice of work as an endogenous variable. Roy [1952] developed the basic framework for such a model of occupational choice, and an important empirical application treating higher education as a choice variable in the estimation of earnings functions can be found in Rosen and Willis [1979]. The basic idea is that individuals self-select in either employment or self-employment on the basis of the gain that might accrue to them. Thus the wage equation may be written as

$$\ell n\, y_i = X_i\beta_i + U_i, \qquad i = E,S, \tag{7}$$

where E stands for employment, S stands for self-employment, y is the wage rate, X is a set of exogenous regressors or individual characteristics, β is the coefficient vector, and U is the random disturbance term. A new variable I, which measures the relative gain of employment over self-employment, can now be constructed as:

$$I = \ell ny_E - \ell ny_S = X_E\beta_E - X_S\beta_S + U_E - U_S. \tag{8}$$

Individuals who gain from employment--those who have $I > 0$--will seek employment. Those with $I < 0$ will go into self-employed activity. Our sample contains individuals who have already self-selected based on their relative gain function in their work status choice.

Table 5

DECOMPOSITION OF VARIANCE FOR LOG EARNINGS IN SIX COUNTRIES

	U.S.	U.K.	France	Mor-occo	Singa-pore	Hong Kong	Pakistan Emp.	Pakistan Full Sample
I. Schooling-Experience Model Estimated in Total Sample:[b] $\ln Y = \text{constant} + rS + a_1X - a_2X^2$								
r	.107	.097	.108	.158	.113	.071	.077	.055
R^2	.285	.316	.350	.443	.449	.249	.392	.260
$\sigma^2(\ln Y)$	.680	.436	.546	.650	.373	.349	.569	.532
$R^2\sigma^2(\ln Y)$	.194	.138	.191	.288	.167	.087	.223	.138
$(1-R^2)\sigma^2(\ln Y)$	.486	.298	.355	.362	.206	.262	.346	.394
R^2	.485	.528	.445	.678	.684	.395	.364	.149
$R^2\sigma^2(\ln Y)$	.330	.230	.243	.441	.255	.138	.207	.079
$(1-R^2)\sigma^2(\ln Y)$	.350	.206	.303	.209	.194	.211	.362	.453
II. Schooling Model Estimated in Overtaking Sample: $\ln Y^* = \text{constant} + rS$								
r	.165	.068	.110	.175	.134	.140	.112	.089
R^2	.328	.105	.340	.696	.596	.444	.461	.294
$\sigma^2(\ln Y^*)$	.520	.230	.459	.687	.400	.379	.667	.641

SOURCES: U.S. from Mincer [1974]; U.K. from Psacharopoulos and Layard [1979]; France from Riboud [1977]; Morocco from Psacharopoulos [1977]; Singapore from Liu and Wong [1980]; Hong Kong from Wong [1981].

[a]Hong Kong and Pakistan estimates use monthly earnings, Singapore estimates use hourly wages, and the rest use annual earnings.

[b]$\ln Y$ denotes log monthly earnings of men in the total sample.

[c]$\ln Y^*$ denotes log monthly earnings of men in the overtaking experience years.

The Rawalpindi results are compared with those of some other countries in Table 5. Both the employee subsample and the full sample results are presented. The results especially for the employee subsample compare well with other countries. Estimates of γ, the return to schooling, are generally lower than in other countries. The $\hat{R}^2$ for the employee sample is roughly comparable to that of other countries while the $\hat{R}^2$ for the full sample is quite low, indicating once again the lack of adequate explanatory variables for the self-employed. The $\hat{R}^2$ presented in Tables 1 and 3 is the Mincer measure of estimating the explanatory powers of schooling in the human capital model. It is defined as

$$\hat{R}^2 = 1 - \frac{\text{Var u in overtaking set}}{\text{Var y in full sample}}$$

For our sample, 15 percent of the variation in earnings is explained by schooling alone; whereas for employees the measure is 26.4 percent. The $\hat{R}^2$ is lower than the regression $\hat{R}^2$, a consequence of the variance of ℓny being larger in the overtaking years than in the whole sample. For Britain, France, and the United States, the $\hat{R}^2$ is larger than the $\hat{R}^2$ in the overtaking years regressions; the opposite results hold for Hong Kong, Morocco, and now Pakistan. Except for Singapore, the less developed countries differ in this respect from the developed countries. An explanation for such a result probably lies in the dynamic effects embodied in cross-section data. Higher income growth rates in the poorer countries coupled with rapid industrialization and technology transfers have in all probability been shifting the demand for skilled or educated manpower outward. Consequently both the demand for schooling and school enrollments would have tended to rise over time to the benefit of younger cohorts. The resulting observed variation in earnings would therefore be reduced, even though lifetime variation, which by construction is what the overtaking set seeks to approximate, is increased.

Table 4

LOG EARNINGS OF ALL MEN
(-VALUES IN PARENTHESES)

Variables	Full Sample		Employees		Self-Employed	
Constant	5.08	5.15	4.80	4.89	5.67	5.69
	(148.86)	(150.37)	(125.15)	(125.44)	(87.54)	(87.93)
S	.055	-.019	.077	.002	.024	-.023
	(20.07)	(-2.32)	(24.64)	(.212)	(4.69)	(-1.48)
S^2		.006		.006		.005
		(9.60)		(8.43)		(3.23)
		[.051]		[.085]		[.017]
EXP	.063	.064	.069	.070	.043	.044
	(20.77)	(21.73)	(17.89)	(18.56)	(8.38)	(8.69)
EXP^2	-.001	-.001	-.001	-.001	-.0008	-.0008
	(-15.32)	(-16.19)	(-12.52)	(-13.00)	(-7.79)	(-8.10)
	[.475]	[.494]	[.497]	[.488]	[.246]	[.235]
	[19.0]	[19.01]	[24.24]	[25.40]	[8.27]	[7.25]
R^2	.260	.287	.392	.418	.088	.099
Obs.	2484	2484	1599	1599	885	885

[a]Values in square brackets under S^2 are estimates of r at the mean value of S.

[b]Values in square brackets under EXP^2 are estimates of k_0 and T respectively.

employed are somewhat surprising and different from results from Hong Kong. Wong [1981] found little to distinguish between the self-employed and the employed with similar estimates for the returns to schooling, γ, initial investment in on-the-job training, k_0, and the time of termination of on-the-job training, T. The Mincer measure of the explanatory power of schooling, $\hat{R}^2$, was different for the two categories even in Hong Kong, although the difference was not as large as in Rawalpindi.

It follows, therefore, and it is further substantiated by the insignificant schooling coefficients in the estimations for the self-employed, that higher levels of schooling yield a significantly larger return in employment than in self-employment.

Earnings function estimates for both the work statuses and for all individuals are presented in Table 4. As expected, estimates of γ, the returns to schooling, are lower than those for the overtaking sets. Nonlinear returns to schooling are once again indicated. The introduction of the quadratic schooling term raises the R^2 but leaves the experience coefficients unaffected. Experience variables are significant with the expected signs indicative of the usual concave age-earnings profile. The experience variable here is an observed variable and not one derived from the difference between age and schooling.

As in the overtaking set, the full sample regressions indicate a small estimate for the returns to schooling among the self-employed. In the sample they are all owner-operators and expected to hold small amounts of capital. The low estimate of returns to schooling is most likely a consequence of our lack of information on capital holdings. In the earnings function estimates for the employed, wages are the dependent variable, whereas for the self-employed the dependent variable is wages plus capital income. To the extent that capital endowments can be substituted for human capital acquisitions (by means of bequests, for example--see Becker [1981]), our estimates of γ for the self-employed will be downward biased.

Table 4 presents estimates of k_0, the initial net investment ratio, and of T (= $k_{0/b}$), the terminal period of on-the-job investments. Lower k_0 and T are indicated for the self-employed. These individuals invest less of their potential earnings on learning on the job, and they finish this learning much earlier than their counterparts in employment. Because the self-employed are an older group of people, our results may merely be capturing vintage and cohort effects. Overall the human capital model fits the employee category well but not that of the self-employed.[6] The observed differences between the self-employed and the

[6] A comparison of the full sample regression coefficients with those of the two work status choices shows that the employee and the full sample results agree. This group therefore dominates in the full sample results.

family earnings than in HH earnings contrasts with the results for the United Kingdom, the United States, and Israel.

Since summary measures are incapable of ranking distributions unambiguously, little can be said about the shape of such distributions from the study of such measures alone. A method for comparing distributions used extensively by Smith and Welch [1979] is presented in Figs. 1 and 2. The earnings of the employed at selected percentiles of the employed earnings distribution are presented relative to the self-employed at the same percentile of the self-employed earnings distribution. For HH earnings, the U-shaped nature of the curve indicates that up to the bottom of the trough, earnings for the self-employed are more unequally distributed. From there on, the upward slope suggests the opposite: i.e., from the fourth to the ninth decile the employed distribution of earnings is more dispersed. The family earnings curve has, however, a more marked positive slope from the third decile onward. Given that a large part of both curves is positively sloped and this positive slope is in the upper deciles, the distributions of both the HH earnings and the family earnings are more skewed for the employed category than for the self-employed. Earnings ratios for family and HH earnings are plotted for families with secondary earners only in Fig. 2. The HH curve follows much the same pattern as before, but family incomes appear to be similarly distributed across employment categories. Contrasting the slope of the two curves in Fig. 1, one can study the effect of the secondary earner earnings on family income. At lower percentiles of the income distribution, secondary earner earnings do tend to decrease relative dispersion, while the opposite holds at the upper end of the distribution.

This finding--that employed earnings are more evenly distributed than self-employed earnings at the lower end of distribution of incomes--is in keeping with the conjecture of labor market segmentation. Earlier it was claimed that the employed/self-employed distinction approximated the dual labor markets of that conjecture. It would be fair to point out, however, that this approximation would increase in accuracy as the observed sample was weighted toward lower-income groups. Alternatively, the informal sector at the lower end of the income scale would comprise

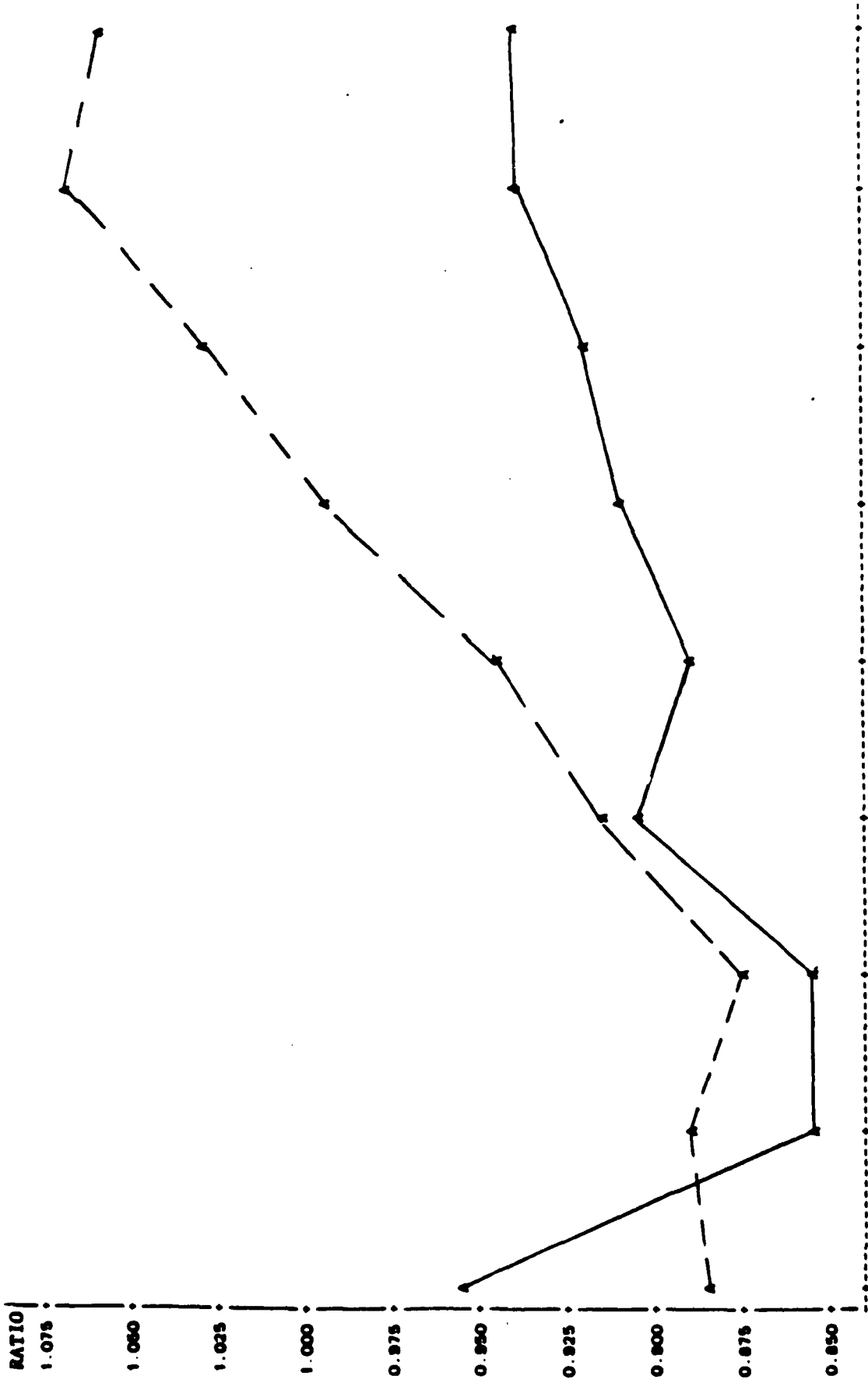


Fig. 1--Ratios of Employed to Self-Employed Incomes for All Families

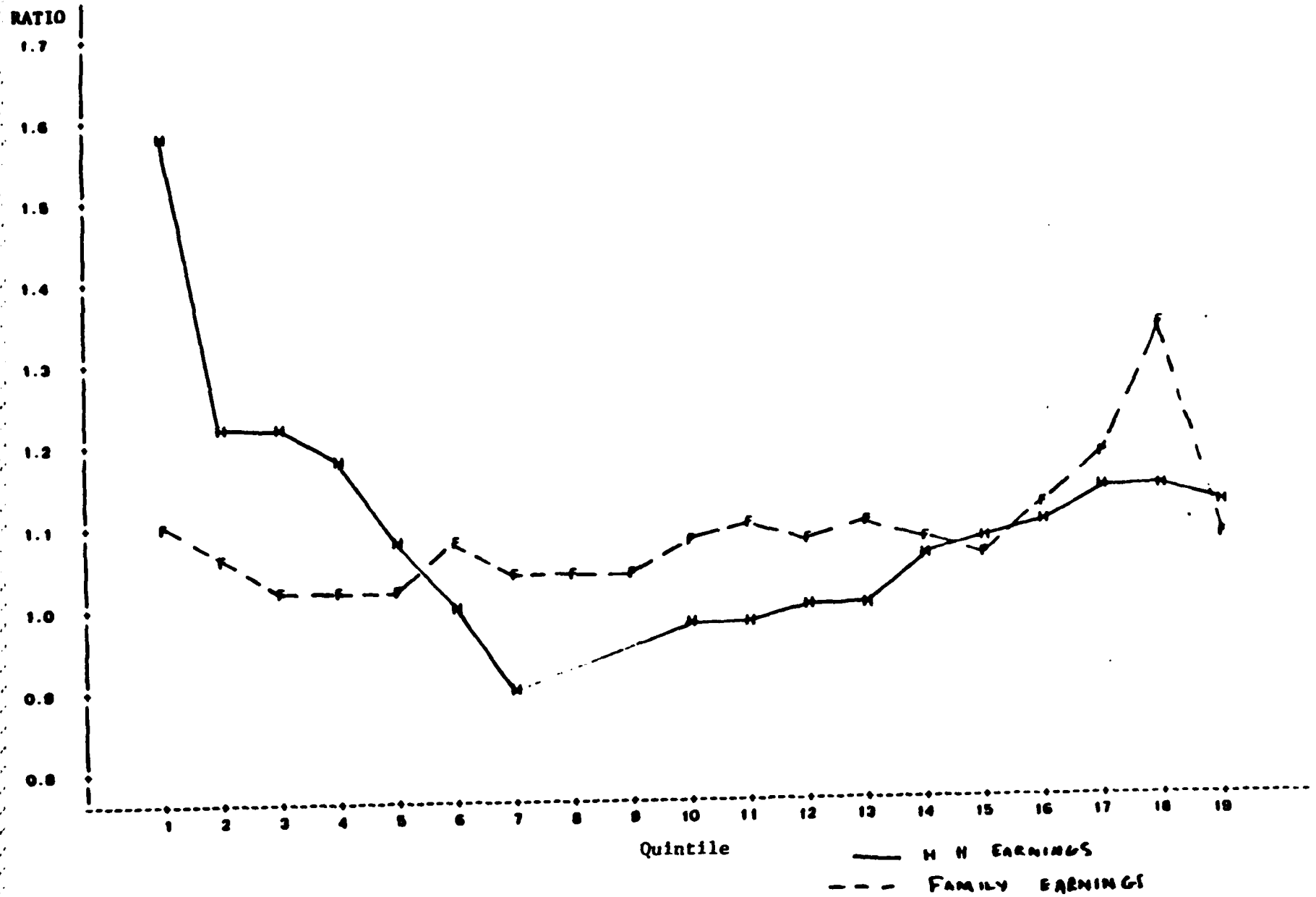


Fig. 2--Ratios of Employed to Self-Employed Incomes for Families with Secondary Workers

mainly the self-employed. The dual-market conjecture claims that more uncertainty would characterize the informal sector incomes. Wage contracts, unionization, and job tenure rules all contribute to reducing inequality in the formal (employed) sector. Whereas the segmentation conjecture is supported by the fact that individual incomes are more unequally distributed in the informal (self-employed) sector, Figs. 1 and 2 show how joint optimization at the family levels leads to relatively greater equality for family incomes.

Further study of the relationship between earnings of workers in a household shows (see Table 9) that incomes of earners in households are positively correlated for the HH employed category for all age groups. To some extent this positive correlation explains the more dispersed family earnings variation for the HH employed group. The large positive correlations at both ends of the age distribution are probably a consequence of many observed two-earner families reporting positive labor supply for both fathers and sons. Self-employed families show workers' earnings to be negatively correlated, revealing the usual compensating role for the secondary earners that has been observed for wives in Western economies (U.S., U.K., and Israel). When stratified by HH education, the self-employed sample shows negative correlations between earners' earnings, whereas the employed sample shows more skewness because of this positive correlation.

V. FAMILY EARNINGS AND SECONDARY EARNERS

Based on their individual evaluations of time, family members choose between market and non-market activities, and thus select their mode of supplementing family incomes. Individual valuations of time in alternative activities, however, might depend on family circumstances in particular HH income. If, with an increase in HH income, the value of nonmarket activity rises by more than the value of market activity, then secondary earner participation would be negatively correlated with HH earnings and observed family earnings would be more equally distributed than true family earnings. In such a case, increases in the correlation between earnings of earners in a family would lead to greater potential earnings distribution relative to the observed family income

Table 9

CORRELATION COEFFICIENTS IN EARNER EARNINGS

A. Correlation coefficients in earner earnings

	Age of HH (years)			
Employed	21-30	31-40	41-50	51-60
All families	.212	.053	.016	.209
Secondary earner families	.365	.116	.364	.216
Self-Employed				
All families	.045	-.175	-.091	-.216
Secondary earner families	.313	-.385	-.046	.148

B. Correlation coefficients and coefficients of variation

	Education of HH											
	Less than 6 years			6 to 10 years			Over 10 years			TOTAL		
Employed	C	C		C	C		C	C		C	C	
All families	.007	.482	.660	.118	.558	.641	-.062	.545	.564	.064	.618	.684
Self-Employed												
All Families	-.126	.651	.661	-.048	.662	.703	-.325	.560	.525	-.117	.670	.658

distribution. Positive transmission of market wage-increasing attributes such as human capital, family influence, and good will, both across and within generations, causes the correlation in earnings of a family's earners to increase. The Becker [1981] analysis of a positive relationship between inherited human capital and incomes of parents would indicate that the observed correlation should be lower than the potential correlation. For our sample this would be the case where earners are mainly sons and brothers of HH. Thus the observed distribution of earnings will be less skewed than the potential distribution that would be observable if all potential earners were in the market.

Since wages for secondary earners are observable only for market participants, censoring corrections will have to be made to obtain the true correlation of earners' earnings. The problem of censoring corrections has been extensively studied in the female labor supply literature (see Heckman [1976] and Smith [1979]). In these analyses the problem normally is the estimation of an earnings function or labor supply schedule for wives, where only some of them report positive earnings and hours worked. Our sample offers us no natural grouping for secondary earners such as wives. Few wives work, whereas a number of other male members do report market earnings. For our analysis, therefore, wives are excluded and potential earners are defined to be male members other than the HH.

The censoring framework is fairly similar to the occupation-choice model discussed earlier. Three wage relationships are specified: the observed HH earnings function,

$$W^{HH} = \alpha X + U_1, \tag{12}$$

the market earnings function for the secondary earner,

$$W^{M} = \beta Y + U_2, \tag{13}$$

and the reservation wage equation for the secondary earner,

$$W^R = CZ + U_3, \tag{14}$$

where W^{HH}, W^M, and W^R are HH market wage, secondary-earner market wage, and secondary-earner reservation wage, respectively. X, Y and Z are exogenous regressors in the respective wage equations, and the U_i's (i = 1,2,3) are the usual disturbance terms. The secondary earner will choose market work only if the market wage is greater than the reservation wage. Defining

$$I = \frac{(\beta Y - CZ)}{\sigma_p} \tag{15}$$

the secondary earner reports a positive wage only if

$$I > \frac{U_3 - U_2}{\sigma_p} \tag{16}$$

where σ_p is the variance of $U_p = U_3 - U_2$. Thus, as stated earlier, the censoring problem may be viewed as one of specification bias, since

$$E(W^S \; 6_p > U_p) = \beta Y + \frac{\sigma_{22} + \sigma_{23}}{\sigma_p} \lambda = \beta Y + \gamma_2 \lambda, \tag{17}$$

where λ is the inverse of the Mill's ratio and is equal to $\frac{f(\hat{I})}{F(\hat{I})}$, the f and F stand for the density and distribution function of the standardized normal variable, and the circumflex(^) denotes the estimated values of I. Similarly, for the HH in families where there are secondary earners,

$$E(W^{HH} \quad \sigma_p > U_p) = \alpha X + \frac{\sigma_{12} - \sigma_{13}}{\sigma_p} \lambda = \alpha X + \gamma_1 \lambda. \tag{18}$$

To correct for censoring bias in residual variances, estimates from Eqs. (17) and (18) can be used. Thus

$$COV(U_i U_j) \; I\sigma_p > U_p) = \sigma_{ij} - \gamma_i \gamma_j (\lambda^2 + I\lambda), \tag{19}$$

where σ_{ij} is the true covariance over the complete sample. The error structure may thus be retrieved by estimating the censoring corrected regressions for both HH and the secondary earner.

A small proportion of observed households, as noted above, reports more than two secondary earners. To attempt to estimate the complete family covariance structure not only would be econometrically intractable but would require a much richer data set than the one we have. The analysis here will therefore be confined to the eldest secondary earner, where there are more than one. For families with no secondary earners, the eldest member other than the wife and HH is regarded as the potential secondary earner. Consequently, unlike the problem of female labor supply, we do not have a homogeneous group of individuals self-selecting themselves into the market. Thus, whereas in the standard problem, women are choosing between housework and market work, the individuals under study here may have more diverse aims. Since we are dealing with sons, siblings, and parents of the principal breadwinner in the family, the likely choice would be between staying at home and accumulating human capital either in the form of schooling (say, for sons) or in the form of direct market experience.

Retention of the sample division by HH employment status choice may be worthwhile in observing behavioral differences in these two important markets. Earnings functions corrected for selection bias are presented in Table 10. The effects of selectivity cannot be rejected, as significant coefficients for γ_2 indicate. Selectivity, however, seems to be weaker in the self-employed subsample. The human capital variables, education, and on-the-job experience are all significant for the secondary earners in both subsamples as well as in the full sample. Wealth effect as measured by HH home ownership in Rawalpindi is now strongly positive, contributing an average of 15 percent to the earnings of a secondary earner.

Work status choice of the secondary earner in the earnings function now gives us an idea of the extent of within-family transmission of work status knowledge and skills. A natural hypothesis would be that a positive correlation between work status choices across generations should translate itself into some market advantage. Thus we observe that a choice of self-employment by the secondary earner in an employed HH household lowers earnings on average by 8.7 percent. On the other hand, a choice of self-employment in the self-employed HH household

Table 10

SELECTIVITY CORRECTED EARNINGS FUNCTION ESTIMATES FOR SECONDARY EARNERS
(-RATIOS IN PARENTHESES)

	Employed HH	Self-employed HH	All
GRAD	.0553	.0610	.0576
	(8.9588)	(6.1601)	(11.3200
YRSWKED	.0447	.0471	.0390
	(4.7459)	(3.9765)	(5.5888)
YR2	-.0009	-.0007	-.0007
	(4.3415)	(3.3591)	(4.9457)
YRSATJOB	.0063	.0064	.0095
	(1.1427)	(.8560)	(2.1612)
CURHSOWN	.1709	.1445	.1559
	(3.1064)	(1.6924)	(3.3396)
E	-.0865	.2647	.0919
	(1.2723)	(3.0721)	(1.7554)
γ_2	.1265	.1003	.1191
	(1.9243)	(1.4009)	(2.5245)
Constant	4.5442	4.4969	4.5356
	(16.3231)	(13.7387)	(21.9914)
R^2	.3177	.2557	.2697
γ_1 [a]	1.3005	1.4765	1.2620

[a]This is the coefficient estimate for the γ_1 from the HH wage regression.

hand, a choice of self-employment in the self-employed HH household increases earnings by 26.5 percent.

The estimates of the coefficients of γ_2 (the selectivity variable) in the secondary-earner earnings functions are all positive. The implication is that in a regression of residuals of the reservation wage equation on the residuals of the secondary-earner market wage equation, an increase of 1 percent in the wage of the secondary earner will produce less than 1 percent increase in reservation wage. Thus, increases in market wages induce less than a commensurate rise in the reservation wage. It follows, therefore, that secondary earners with the highest market wage opportunities will be in the market sector. Such an intuitively plausible result is strengthened when it is taken into consideration that the opportunity cost of market time is not house work for the individuals under study, since the wife is present for such work in all households under consideration.

It is likewise intuitively plausible to expect that the unobservables in the market wage of the HH and the secondary earner will be more highly correlated than the unobservables in the market wage of HH and the unobservables of the secondary-earner reservation wage. The positive coefficients of γ_1 in the HH earnings functions imply such a result, and thus confirm the hypothesis of positive within-family transmission of human capital and other earnings-enhancing characteristics.

As mentioned earlier, the observed variance in family earnings should be expected to be smaller than the potential distribution, when all potential earners are in the market. Using the censoring corrected wage regressions, the true covariance structure of family earnings is now computed and presented in Sec. A of Table 11. As expected, censoring corrections increase both the residual variance for the secondary-earner earnings and the covariance between the earnings of earners in a household. Censoring corrections seem to be more important for the self-employed, since both the covariance and the correlation in the earnings within a household are increased by more as a result of corrections for this group. The larger correlation between earners in the employed subsample may explain, in part, our earlier observation of a more dispersed family income distribution for the employed.

Earnings were predicted using the censored wage equations for all potential secondary earners. Section B of Table 11 presents the correlation coefficients for the residual, predicted, and predicted-plus-residual earnings for HH and the secondary earners. Once again the correlation in earnings is larger in the employed families; this is not a consequence of the larger residual correlation but of the larger correlation in predicted earnings. Censoring corrections would therefore indicate that potential family earnings are more dispersed than observed earnings. We also observe that even after censoring corrections, employed family earnings are more unequally distributed than self-employed family earnings.

Table 11

COVARIANCES AND CORRELATIONS IN FAMILY EARNINGS

A. Estimated Residual Variances in Earnings Equations

Employed	Censored	Full
σ_{11}	.0579	1.6640
σ_{22}	.2764	.2916
σ_{12}	.0211	.1774
ρ_{12}	.1668	.2547
Self-Employed		
σ_{11}	.0655	2.1340
σ_{22}	.4061	.4156
σ_{12}	.0028	.1433
ρ_{12}	.0172	.1522
Full Sample		
σ_{11}	.0838	1.5958
σ_{22}	.3347	.3482
σ_{12}	.0132	.1559
ρ_{12}	.0785	.2092

B. Correlation and Variances in Earnings

Correlation	*Employed*	*Self-Employed*	*All*
Residual earnings	.2547	.1522	.2092
Predicted earnings	.5071	.1409	.3793
Total correlation	.2301	.1223	.2400
Variances			
HH	.4987	.5725	
Secondary	.4208	.5558	.4722
Censored secondary earner	.4060	.5477	.4590

VI. CHOICE OF FAMILY STRUCTURE AND EARNINGS

Since observed earnings are less widely dispersed than potential earnings, and since this result holds across HH work status choice, the principal effect of secondary-earner labor force participation, therefore, seems to be the equalizing of observed or money incomes

across households. The secondary earners are, however, not the ones usually observed in developed economies, the wives of households, but male members such as sons and siblings of HH. It is possible, therefore, that our results are indicative of the choice of family structure. Poorer families may group together to jointly consume "household" goods more intensively and to enable more members to be freed from the household to increase the grouping's market power. Standardizing for family grouping, therefore, will produce an even more greatly skewed distribution of income. Likewise, for richer groups the likelihood of working sons and siblings living away and thus beyond observation of our sample would be greater. Once again, choice of family structure influences the distribution of observed family incomes.

To obtain some understanding of the relationship between choice of family structure, family wealth structure, and secondary-earner labor force participation, probit functions were estimated for the probability of having one or more secondary earners in the household. These estimates are presented in Table 12. Variables in the probit include HH education, earnings, and employment status. Also included are family wealth variables and a variable that identifies an "extended" family. An extended family is defined as one that is not nuclear, the latter being one that includes a wife and children under 20 years of age. The extended family variable is, as expected, the most important variable in increasing the probability of a household having a secondary earner in its fold. Also, in keeping with expectations, is the finding that an increase in HH earnings lowers the probability of finding a secondary earner in the household. Ownership of residence by HH also works in the same direction as HH income. Thus the wealth effect has a negative effect on the probability of a secondary earner in a household. HH education is insignificant, while self-employment of HH has a negative effect, possibly because of family participation in the HH family business.

These results are reinforced when potential earners are identified from each family and the effects of the determinants of the probability of labor force participation of these individuals is estimated. The results of such estimations are presented in Table 13. The significant and negative coefficients for the income of the principal breadwinner

Table 12

MAXIMUM LIKELIHOOD ESTIMATES OF THE PROBABILITY OF HAVING MORE THAN ONE EARNER IN A HOUSEHOLD

	1 if more than one earner	1 if more than two earners
lny	-.5477	-.5003
	(6.100)	(4.7010)
	[.0005]	[.0013]
XTEN	1.0758	-.9929
	(12.042)	(7.9033)
	[.3992]	[.3355]
HEDGRAD	-.0114	.0052
	(-.9743)	(.3452)
	[0.45]	[.0021]
HEDAGE	.0256	.0156
	(6.326)	(3.2416)
	[.0010]	[.0050]
FCLAS	-.0103	-.0332
	(.8433)	(1.9480)
	[-.0041]	[-.0132]
MIG	-.0420	-.4404
	(.3054)	(2.1080)
	[-.0168]	[-.1754]
HE	-.3094	-.2259
	(3.225)	(1.8681)
	[-.1223]	[-.0900]
CURPUCCA	-.0609	-.1875
	(.4038)	(1.0483)
	[.0243]	[.0737]
CURHSOWN	-.1800	[1048]
	(2.027)	(.9378)
	[.3974]	[.0510]
Constant	1.6573	1.1281
	(2.916)	(1.6281)
-2 log L	595.561	342.387
No. of multiple-earner households:	392	127

Note: Parentheses indicate asymptotic t-ratios, square brackets the marginal probability at the means.

and his education suggest that relatively wealthier families are less likely to have secondary earners living in the household. The independent significant and negative effect of the principal breadwinner's education is, in all likelihood, indicative of a household

Table 13

MAXIMUM LIKELIHOOD ESTIMATES OF THE PROBABILITY OF LABOR FORCE PARTICIPATION BY SECONDARY EARNERS

	Coefficient	Asymptotic t-ratio
XTEN	.2024	2.4017
AGE	.0041	1.3885
HEDAGE	.0015	4.5161
HEDLNY	-.0038	5.9167
HEDGRAD	-.0274	3.0580
HE	-.3376	-4.1839
Constant	1.6397	
-2 Log L	816.306	

preference for education in a choice between market work and education, holding household wealth constant. The age of the head of household has a positive coefficient and is probably a consequence of the fact that an older man is more likely to have working dependents living in. Individual's own age when family background is controlled for is not relevant to the decision to participate. The self-employed employment status of the head of household is significant and negative in affecting the work decision. Finally, extension of the family is an important variable with a positive effect on the work decision, indicating once again the economic role of family formation. Extended families, therefore, are formed by the poorer income group for, say, the exploitation of the economies of scale of joint living. Moreover, these extended families enable the release of more workers into the market.

The conclusion seems to be that secondary earners are found in families where a need exists to supplement family earnings. The secondary earners are also more likely to be extended family members, as defined here. An improvement in the economic conditions of the family--in the form, say, of increased earnings of either the HH or the secondary earner--would move the family into the direction of becoming more nuclear and of being supported purely by HH earnings.

VII. CONCLUSIONS

A major concern of development policy has been the achievement of a degree of equity in the distribution of incomes and welfare. An understanding of the determinants of these distributions is, therefore, of some importance to the framing of such policy. In this spirit, the results of the above analysis can be considered to be quite encouraging. It was determined that the human capital model is useful in explaining the distribution of earnings on Pakistan, as elsewhere. Estimated returns to education, especially those for the employed, compare favorably with similar estimates from other countries. However, returns to schooling for the self-employed, a group which constitutes about a third of our sample, remain very weak, even when the endogeneity of the individual work-status decision is incorporated in the estimation procedure. To an extent, an explanation could lie in a combination of a measurement error in reported income because of the inability to measure capital holdings, and the vintage effect in the self-employed education, since they are an older group. Nevertheless, the framing of education policy, which in Pakistan is based so heavily on subsidies, might find the returns estimated here and the international comparison of these returns relevant.

When studying the economic implications of the individual work-status choice, it was seen that the self-employed are earning more than if they had not been able to make their choice. This conforms with the hypothesis of individuals self-selecting on the basis of comparative advantage. A somewhat more puzzling result and one which does not conform to this hypothesis is that the employed are not earning as much as they would have, had they been unable to choose freely, being randomly assigned to a work-status. Although the possibility of some form of labor market imperfections seems to exist, the analysis does provide some contradictory results to a hypothesis from the development literature on labor market segmentation. Contrary to this hypothesis, the self-employed as a group seem to be doing better in the city than the employed. Consequently there is little reason to expect that the formal employed sector would seek to protect itself by means of entry barriers.

In studying the distribution of income, we discover that a finding of a more skewed distribution of income for the employed HH group runs counter to the segmentation hypothesis. However, among the poorer income groups where it can be expected that the formal/informal dichotomy of the segmentation hypothesis conforms better to our categorization of employed/self-employed, inequality in measured earnings is larger for the self-employed. Consequently a more conclusive analysis of the segmentation hypothesis would need a richer dataset.

The role of the secondary earner is similar to that of the wife in developed economies--that of equalizing observed family incomes. However, in our sample, since female labor force participation is so low, the observed secondary earners tend to be male extended-family earners whose opportunity cost of not working is not housework. Observed labor supply behavior at the family level, again, tends to be similar to that in the developed economies, with poorer families putting out more earners. Insofar as for some of these poorer families, the secondary earners are children who, in order to supplement family income have had to forego an education, inequities may persist or even amplify over the generations. Thus, though the economic role of the choice of family structure seems to be to allow the poor to exploit the economies of scale of joint living, enabling more workers per family to be released for wage labor, the important consequences for the dynamics of socio-economic mobility in the underlying population need further analysis.

REFERENCES

Becker, G. S., *Human Capital*, New York: Columbia University Press, 1964.

Becker, G. S., *A Treatise on the Family*, Harvard University Press, 1981.

Chiswick, C. U., "On Estimating Earnings Functions for LDCs," *Journal of Development Economics* 3 (March 1977): 67-78.

Gronau, R., "Wives' Labour Force Participation, Differential and Family Income Inequality: The Israeli Experience," *NBER Working Paper* No. 668, April 1981.

Hamdani, K., "Education and the Income Differentials," *Pakistan Development Review*, Summer 1977, Vol. 46, No. 2 (Summer 1977): 144-164.

Haque, N., "An Economic Analysis of Personal Earnings in Rawalpindi City," *Pakistan Development Review*, Vol. 46, No. 4 (Winter 1977): 353-382.

Heckman, J.J., "Sample Selection Bias as a Specification Error," in *Female Labor Supply: Theory and Estimation*, edited by J. P. Smith, Princeton, New Jersey: Princeton University Press, 1979.

Layard, R., and A. Zabalza, "Family Income Distribution: Explanation and Policy Evaluation," *Journal of Political Economy*, Vol. 87, No. 5, Part 2 (1979): 5133-5161.

Lee, L. F., "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables," *International Economic Review*, Vol. 19, No. 2 (June 1978): 415-433.

Liu, P. W., and Y. C. Wong, "Human Capital and Inequality in Singapore," *Economic Development and Cultural Change*.

Muzumdar, D., *The Urban Labor Market and Income Distribution: A Study of Malaysia*, Oxford University Press, 1981.

Mincer, J., *Schooling Experience and Earnings*, New York: Columbia University Press, 1974.

Psacharopoulos, G., "Schooling Experience and E[illegible]ngs: The Case of an LDC," *Journal of Development Economics* 4 (March 1977): 39-48.

Psacharopoulos, G., and R. Layard, "Human Capital and Earnings: British Evidence and a Critique," *Review of Economic Studies* 46 (July 1979): 485-503.

Riboud, M., "An Economic Analysis of Earnings Distribution in France," Ph.D. dissertation, University of Chicago, 1977.

Rosen, S., and R. J. Willis, "Education and Self-Selection," *Journal of Political Economy*, Vol. 87, No. 5, Part 2 (October 1979): S7-S36.

Roy,. A. D., "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, N.S. 3 (June 1954): 135-46.

Smith, J. P., "The Distribution of Family Earnings," *Journal of Political Economy*, Vol. 87, No. 5, Part 2 (October 1979): S163-S192.

Smith, J. P., and F. Welch, "Inequality: Race Differences in the Distribution of Earnings," *International Economic Review*, Vol. 20, No. 2 (June 1979): 515-526.

Smith, J. P., *Female Labor Supply: Theory and Estimation*, Princeton, New Jersey: Princeton University Press, 1979.

Wong, Y. C., "Family, Work Status and Earnings Distribution in Hong Kong," Ph.D. dissertation, University of Chicago, 1981.

END

FILMED

11-85

DTIC